

Soutenance de Stage

Hao ZHANG

Tuteur de stage: Thibaud HULIN

Enseignant référent du stagiaire: Besma ZEDDINI



Plan

- ❖ Introduction
- ❖ Présentation du laboratoire
- ❖ Présentation du projet HUMANE
- ❖ Présentation du stage
- ❖ Mise en œuvre de stage
- ❖ Conclusion

Introduction

- ❖ **Thème de stage :** Développement de l'ontologie et ses services intelligents
 - ❖ Convertir automatiquement des cartes mentales (il y a déjà la version v0.1) en une ontologie.
 - ❖ Convertir automatiquement des textes collaboratifs en ligne en une ontologie.
 - ❖ Sémantiser les données de bases d'expérience en innovation pédagogique (ÉduBase) pour les relier à l'ontologie et en permettre l'exploitation.
 - ❖ Développer des services intelligents à partir de données collectées à l'aide d'une plateforme sémantique comme Whyis ou Semantic Mediawiki.

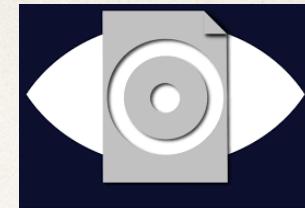
Présentation du laboratoire

- ❖ L'EA 4661 ELLIADD (Edition, Littératures, Langages, Informatique, Arts, Didactiques, Discours) est une unité de recherche de l'Université de Franche-Comté reconnue par le Ministère et évaluée A par l'AERES pour son projet scientifique.
- ❖ ELLIADD regroupe 74 enseignants-chercheurs dont 23 HDR, une centaine de doctorants et 30 chercheurs associés, dans plusieurs champs disciplinaires associés : sciences du langage (CNU-7), langue et littérature françaises (CNU-9), sciences de l'information et de la communication (CNU-71), arts de la scène et musicologie (CNU-18), sciences et techniques des activités physique et sportives (CNU-74) et sciences de l'éducation (CNU-70), mais aussi informatique, langues et littératures anglo-saxonnes, germaniques et slaves, psychologie, mécanique, histoire.

Présentation du projet HUMANE

- ❖ **HUMANE** est un projet de recherche financé par le ministère de l'Éducation nationale (via DNE) pour la période 2020-2022.
- ❖ Co-animateurs scientifiques : Béatrice Drot-Delange (sciences de l'éducation) et **Thibaud Hulin** (Sciences de l'Information et de la Communication).
- ❖ Rattaché au laboratoire **ELLIADD**, ce projet concerne tous ses pôles, 7 chercheurs ELLIADD travaillent déjà au projet HUMANE. Il s'agit donc d'un véritable programme de recherche structurant.

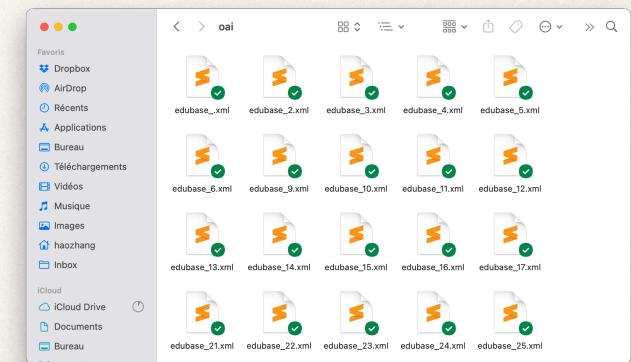
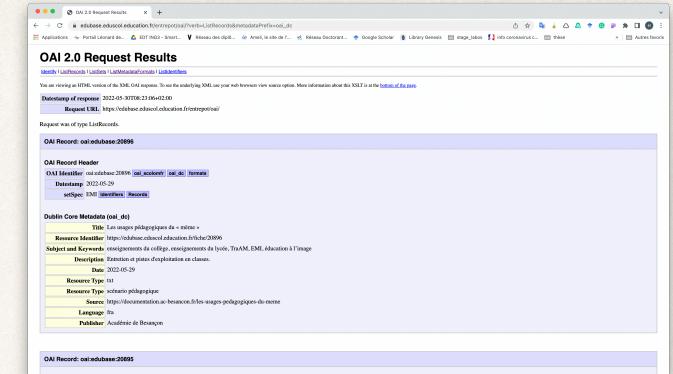
Présentation du stage I



- ❖ OAI-PMH est le sigle de l'Open Archives Initiative - Protocol for Metadata Harvesting, ce qui signifie « protocole pour la collecte de métadonnées de l'Initiative pour les Archives ouvertes ».
- ❖ Le protocole OAI-PMH est un moyen d'échanger sur Internet des métadonnées entre plusieurs institutions, afin de multiplier les accès aux documents numériques.
- ❖ Sickle est une bibliothèque client OAI-PMH légère écrite en Python. Elle a été conçue pour récupérer facilement les données des interfaces OAI à la manière de Python.

```
|: edubase_20912.xml
edubase_20912.xml

1 <ns0:record xmlns:ns0="http://www.openarchives.org/OAI/2.0/" xmlns:ns1="http://ltsc.ieee.org/xsd/LOM
2   xmlns:ns3="http://www.lom-fr.fr/xsd/LOMFR" xmlns:ns4="http://www.lom-fr.fr/xsd/SCOLOMFR" xmlns:nsi
3   ="http://www.w3.org/2001/XMLSchema-instance">
4     <ns0:header>
5       <ns0:identifier oai:edubase_20912></ns0:identifier>
6       <ns0:datestamp>2022-06-01</ns0:datestamp>
7       <ns0:setSpec>ST2S</ns0:setSpec>
8     </ns0:header>
9     <ns0:metadata>
10      <ns1:lom xsi:schemaLocation="http://www.lom-fr.fr/scolomfr/outils/xsd/scolomfrv30/
11        scolomfr.xsd">
12        <ns1:general>
13          <ns1:identifier>
14            <ns1:catalog-URI></ns1:catalog>
15            <ns1:entry>https://ww2.ac-poitiers.fr/biochimie/spip.php?article451</ns1:entry>
16          </ns1:identifier>
17          <ns1:title>
18            <ns1:string language="http://id.loc.gov/vocabulary/iso639-2/fra">TraAM
19              min-spectro DIY et son interface graphique</ns1:string>
20            </ns1:title>
21            <ns1:language>http://id.loc.gov/vocabulary/iso639-2/fra</ns1:language>
22            <ns1:description>
23              <ns1:string language="http://id.loc.gov/vocabulary/iso639-2/fra">Mise à
24              disposition d'un ensemble de ressources en ligne et Playlist de vidéos pour
25              construire un spectrophotomètre (Mini-Spectro) et l'utiliser dans le cadre des
26              sciences expérimentales avec les élèves. Le logiciel libre qui permet de
27              contrôler l'appareil peut aussi être utilisé indépendamment de celui-ci pour un
28              travail en saite informatique ou à la maison.</ns1:string>
29          </ns1:description>
30        </ns1:general>
31        <ns1:lifeCycle>
32          <ns1:status>
33            <ns1:source>SCOLOMFRv2.1.</ns1:source>
34            <ns1:value>http://data.education.fr/voc/scolomfr/concept/final</ns1:value>
35            <ns1:label>final</ns1:label>
36          </ns1:status>
37        </ns1:lifeCycle>
38        <ns1:metaMetadata>
39          <ns1:metadataSchema>Scolomfr v4.0.0</ns1:metadataSchema>
40          <ns1:language>
41            <ns1:source>SCOLOMFRv4.0.</ns1:source>
42            <ns1:value>http://id.loc.gov/vocabulary/iso639-2/fra</ns1:value>
43            <ns1:label>français (fra)</ns1:label>
44          </ns1:language>
45        </ns1:metaMetadata>
46        <ns1:technical>
47          <ns1:location>https://ww2.ac-poitiers.fr/biochimie/spip.php?article451</ns1:location>
48        </ns1:technical>
49        <ns1:educational>
50          <ns1:learningResourceType>
51            <ns1:source>SCOLOMFRv4.0.</ns1:source>
52            <ns1:value>http://data.education.fr/voc/scolomfr/concept/
53              scolomfr-voc-010-num-020</ns1:value>
54            <ns1:label>scénario pédagogique</ns1:label>
55          </ns1:learningResourceType>
56          <ns1:intendedEndUserRole>
57            <ns1:source>SCOLOMFRv4.0.</ns1:source>
58            <ns1:value>http://data.education.fr/voc/scolomfr/concept/teacher</ns1:value>
59            <ns1:label>enseignant</ns1:label>
60          </ns1:intendedEndUserRole>
61          <ns1:context>
```



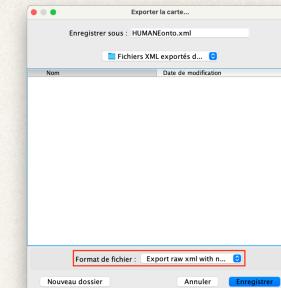
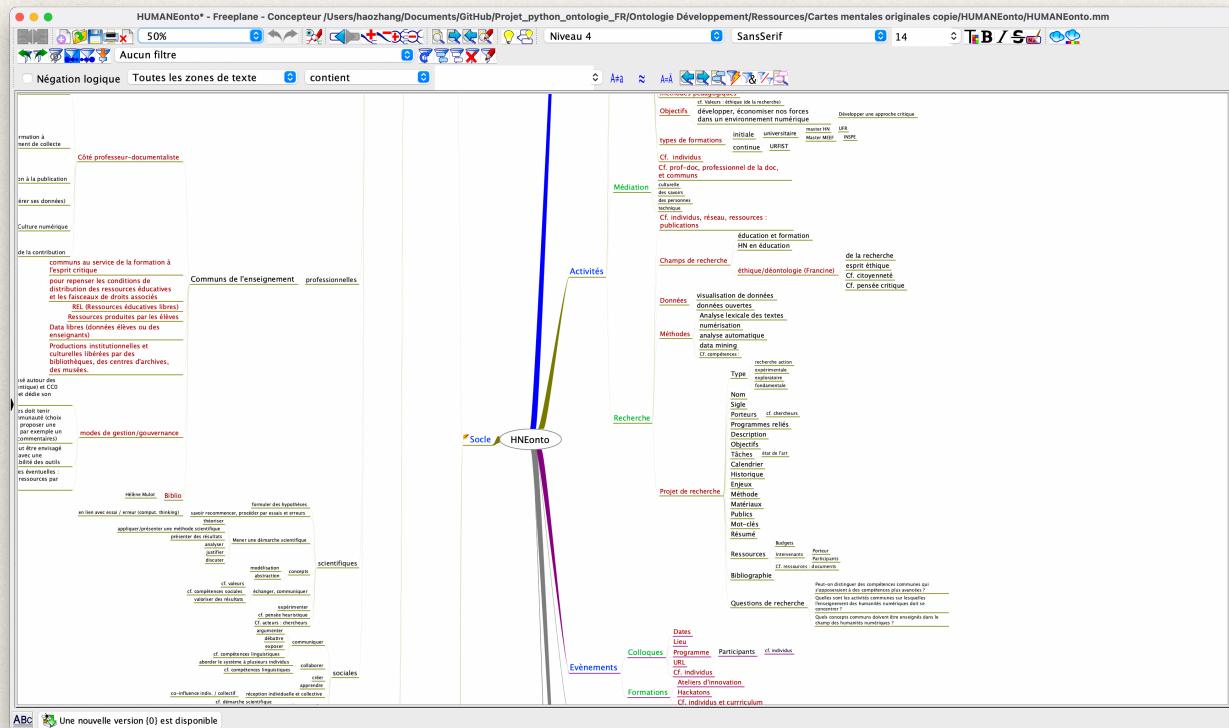
Présentation du stage

III



- ❖ Grâce la fonction d'export de Freeplane, cartes mentales écrites par des experts peuvent être converties en fichiers XML, et le contenu des nœuds peut être lu couche par couche via des programmes Python.
- ❖ La plus grande différence entre la conversion de fichiers HTML et la conversion de fichiers XML est que le premier accorde plus d'attention au contenu, tandis que le second accorde plus d'attention à la structure.

```
<node TEXT="Modélisations" COLOR="#000000" STYLE="fork" NUMBERED="false" FORMAT="STANDARD_FORMAT" TEXT_ALIGN="DEFAULT" MAX_WIDTH="378" MAX_WIDTH_QUANTITY="10 cm" MIN_WIDTH="0" MIN_WIDTH_QUANTITY="0 cm" BORDER_WIDTH_LIKE_EDGE="false" BORDER_WIDTH="1 px" BORDER_COLOR_LIKE_EDGE="true" BORDER_COLOR="#808080" BORDER_DASH_LIKE_EDGE="false" BORDER_DASH="SOLID" POSITION="right" ID="ID_358679034" CREATED="1644577654245" MODIFIED="1644577657282" ICON_SIZE="12 pt">
    <edge STYLE="bezier" COLOR="#808080" WIDTH="1" DASH="SOLID" />
    <font NAME="SansSerif" SIZE="10" BOLD="false" STRIKETHROUGH="false" ITALIC="false" />
</node>
<node TEXT="Socle" COLOR="#0033ff" STYLE="fork" NUMBERED="false" FORMAT="STANDARD_FORMAT" TEXT_ALIGN="DEFAULT" MAX_WIDTH="378" MAX_WIDTH_QUANTITY="10 cm" MIN_WIDTH="0" MIN_WIDTH_QUANTITY="0 cm" BORDER_WIDTH_LIKE_EDGE="false" BORDER_WIDTH="1 px" BORDER_COLOR_LIKE_EDGE="true" BORDER_COLOR="#808080" BORDER_DASH_LIKE_EDGE="false" BORDER_DASH="SOLID" POSITION="left" ID="ID_1439104426" CREATED="161528756353" MODIFIED="1623323639069" ICON_SIZE="12 pt">
    <edge STYLE="sharp_bezier" COLOR="#7c7c00" WIDTH="8" DASH="SOLID" />
    <font NAME="SansSerif" SIZE="18" BOLD="false" STRIKETHROUGH="false" ITALIC="false" />
<richcontent TYPE="NOTE" CONTENT_TYPE="xml"/>
    <html>
        <head>
        </head>
        <body>
            <p> socle plus ou - moins commun ?
            </p>
            <p> à transmettre : impossibilité de transmettre à tout individu de la planète
            </p>
            <p> immatériel
            </p>
        </body>
    </html>
</richcontent>
<node TEXT="Compétences" COLOR="#00b439" STYLE="fork" NUMBERED="false" FORMAT="STANDARD_FORMAT" TEXT_ALIGN="DEFAULT" MAX_WIDTH="378" MAX_WIDTH_QUANTITY="10 cm" MIN_WIDTH="0" MIN_WIDTH_QUANTITY="0 cm" BORDER_WIDTH_LIKE_EDGE="false" BORDER_WIDTH="1 px" BORDER_COLOR_LIKE_EDGE="true" BORDER_COLOR="#808080" BORDER_DASH_LIKE_EDGE="false" BORDER_DASH="SOLID" ID="ID_961912559" CREATED="1610356565961" MODIFIED="1619453228155" ICON_SIZE="12 pt">
    <edge STYLE="bezier" COLOR="#7c7c00" WIDTH="thin" DASH="SOLID" />
    <font NAME="SansSerif" SIZE="16" BOLD="false" STRIKETHROUGH="false" ITALIC="false" />
    <node TEXT="Attributs" COLOR="#990000" STYLE="fork" NUMBERED="false" FORMAT="STANDARD_FORMAT" TEXT_ALIGN="DEFAULT" MAX_WIDTH="378" MAX_WIDTH_QUANTITY="10 cm" MIN_WIDTH="0" MIN_WIDTH_QUANTITY="0 cm" BORDER_WIDTH_LIKE_EDGE="false" BORDER_WIDTH="1 px" BORDER_COLOR_LIKE_EDGE="true" BORDER_COLOR="#808080" BORDER_DASH_LIKE_EDGE="false" BORDER_DASH="SOLID" FOLDED="true" ID="ID_1883853373" CREATED="1617894678060" MODIFIED="1619453281441" ICON_SIZE="12 pt">
        <edge STYLE="bezier" COLOR="#7c7c00" WIDTH="thin" DASH="SOLID" />
        <font NAME="SansSerif" SIZE="14" BOLD="false" STRIKETHROUGH="false" ITALIC="false" />
        <node TEXT="compétences spécifiques aux interactions numériques" COLOR="#000000" STYLE="fork" NUMBERED="false" FORMAT="STANDARD_FORMAT" TEXT_ALIGN="DEFAULT" MAX_WIDTH="378" MAX_WIDTH_QUANTITY="10 cm" MIN_WIDTH="0" MIN_WIDTH_QUANTITY="0 cm" BORDER_WIDTH_LIKE_EDGE="false" BORDER_WIDTH="1 px" BORDER_COLOR_LIKE_EDGE="true" BORDER_COLOR="#808080" BORDER_DASH_LIKE_EDGE="false" BORDER_DASH="SOLID" ID="ID_1313277673" CREATED="1617894681567" MODIFIED="1618848368698" ICON_SIZE="12 pt">
```

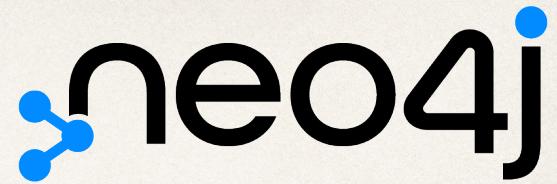


Présentation du stage III



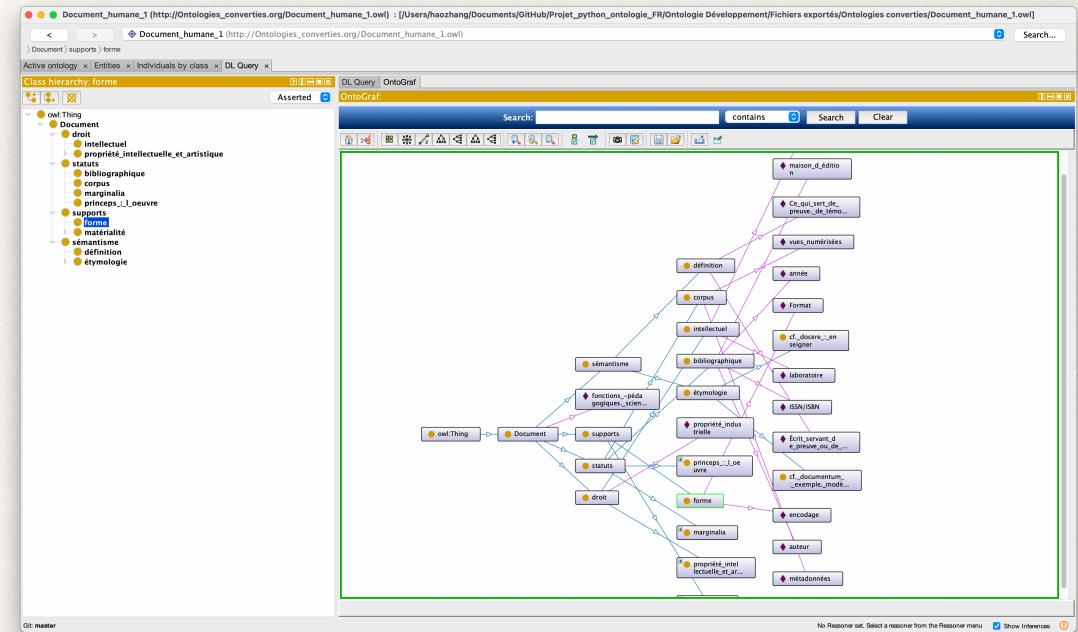
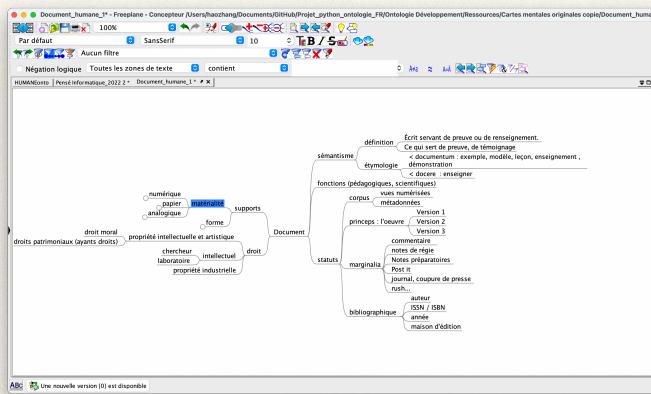
- ❖ Pourquoi ai-je choisi de créer des pages Web via flask / django
 - ❖ L'option 1 (whyis) manque de mises à jour et de support technique correspondant, et nous avons rencontré beaucoup de difficultés lors des étapes d'installation et de test.
 - ❖ L'option 2 (Semantic Mediawiki)les packages associés doivent être mis à jour. En cours d'utilisation, l'espace d'extension est petit et il y a peu de moyens de les adapter en fonction des besoins.
 - ❖ L'option 3 nécessite beaucoup d'expérience et de courage pour investir dans les premières étapes, mais il y a beaucoup de possibilités pour de futurs développements, qui peut être entièrement adaptée en fonction des besoins.

Présentation du stage IV



- ❖ Pourquoi ai-je choisi neo4j
 - ❖ Neo4j offre une fonction de graphe similaire à au fonctionnement des ontologies.
 - ❖ Comparée à la méthode de requête SPARQL couramment utilisée par les ontologies (similaire à la méthode SQL des bases), la méthode de requête de neo4j est plus facile à simuler la méthode de raisonnement d'Ontology.
 - ❖ Et la façon de construire des pages Web à partir de données est une manière courante et stable.
 - ❖ Neo4j lui-même fournit également la fonction de construction d'un système de recommandation, et il est utile de mieux implémenter le système de recommandation via Neo4j lui-même ou en utilisant des technologies avancées.

Mise en œuvre de stage



Database Information

Use database **neo4j**

Node Labels

- (14,425) catalog context
- copyrightAndOtherRestrictions
- cost intendedEndUserRole
- language learningResourceType
- purpose record setSpec
- status taxon

Relationship Types

- (179,584) has_catalog
- has_context
- has_copyrightAndOtherRestrictions
- has_cost
- has_intendedEndUserRole
- has_language
- has_learningResourceType
- has_purpose
- has_setSpec
- has_status
- has_taxon
- has_taxonomyPath

Property Keys

- born catalog_IRI catalog_id
- catalog_identity catalog_label
- Context_IRI context_label
- context_source context_value

neo4j:\$

```
neo4j:$ MATCH (taxon:taxon) RETURN COUNT(taxon) AS taxon
```

taxon

Table	Count
taxon	2627

Started streaming 1 records after 1 ms and completed after 2 ms.

neo4j:\$

```
neo4j:$ MATCH (record:record) RETURN COUNT(record) AS count_record
```

count_record

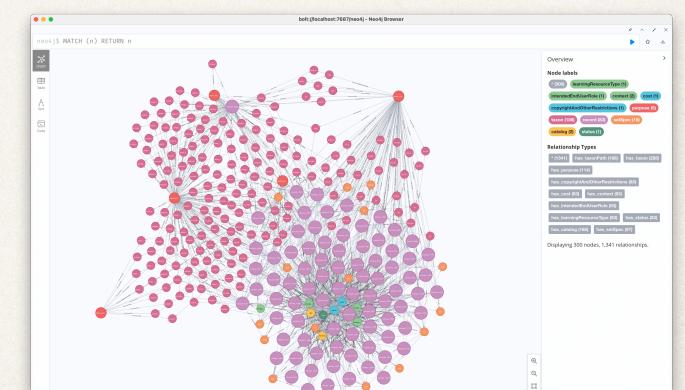
Table	Count
count_record	11730

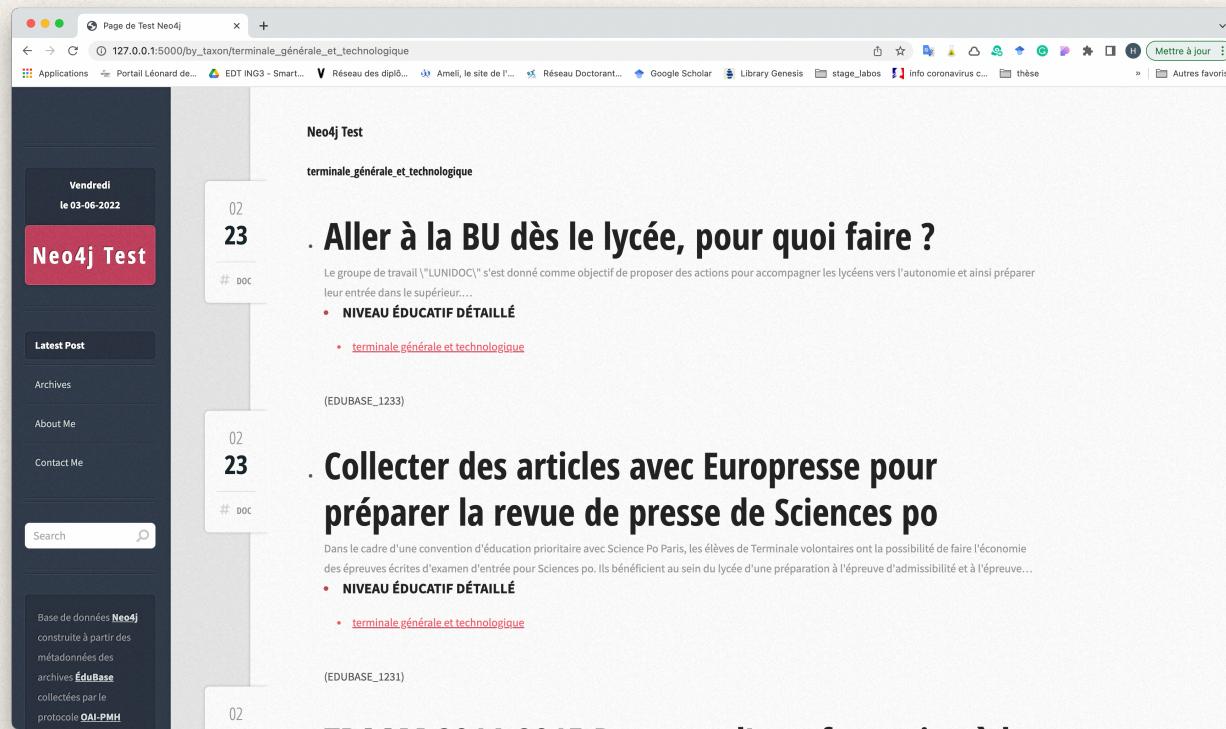
Started streaming 1 records after 3 ms and completed after 3 ms.

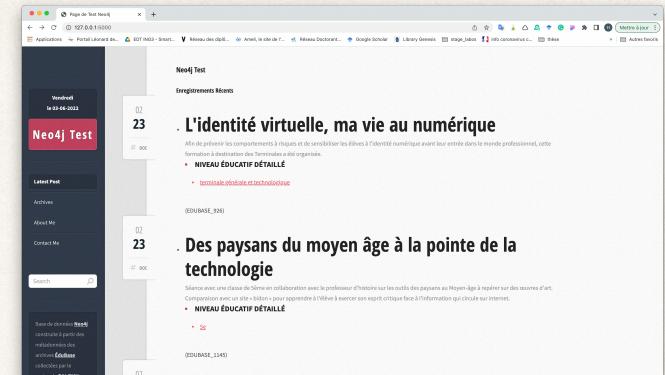
neo4j:\$

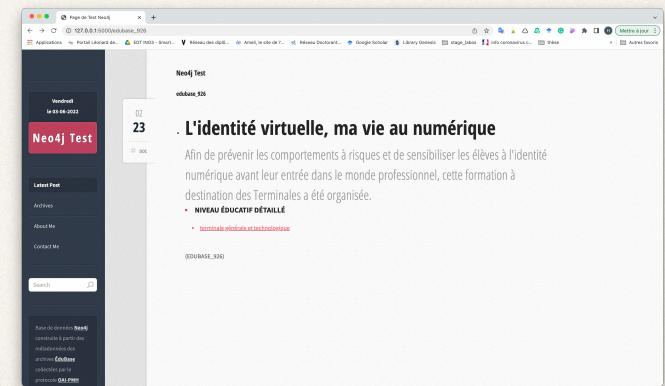
```
neo4j:$ MATCH (n) RETURN n
```

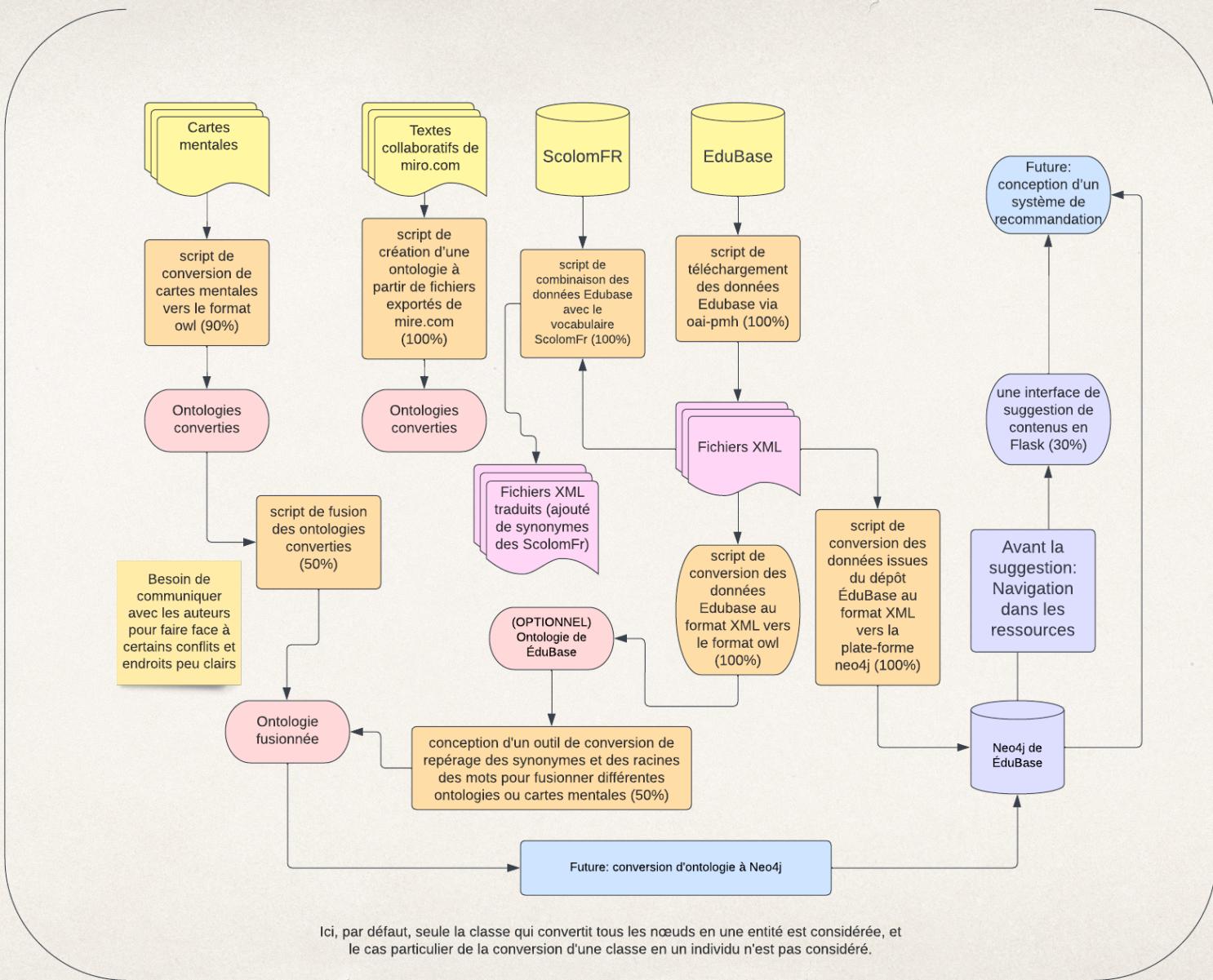
Node Properties



A screenshot of a Neo4j test application interface. The main content area displays a post titled ". Aller à la BU dès le lycée, pour quoi faire ?". Below the title, it says: "Le groupe de travail 'L'UNIDOC' s'est donné comme objectif de proposer des actions pour accompagner les lycéens vers l'autonomie et ainsi préparer leur entrée dans le supérieur...". A bullet point lists ". NIVEAU ÉDUCATIF DÉTAILLÉ" and "terminale générale et technologique". At the bottom of the post, there is a link "(EDUBASE_1233)". On the left side, there is a sidebar with a search bar and a section titled "Base de données Neo4j" which describes the system as being built from metadatabases of archives Edubase collected by the protocol GAI-PMH.

A screenshot of a Neo4j test application interface. The main content area displays a post titled ". Des paysans du moyen âge à la pointe de la technologie". Below the title, it says: "Séance avec une classe de 5ème en collaboration avec le professeur d'histoire sur les outils des paysans au Moyen Âge à repérer sur des œuvres d'art. Compréhension pour ce site à telles + leur apprendre à l'école à exercer son esprit critique face à l'information qui circule sur Internet...". A bullet point lists ". NIVEAU ÉDUCATIF DÉTAILLÉ" and "terminale générale et technologique". At the bottom of the post, there is a link "(EDUBASE_1145)". On the left side, there is a sidebar with a search bar and a section titled "Base de données Neo4j" which describes the system as being built from metadatabases of archives Edubase collected by the protocol GAI-PMH.

A screenshot of a Neo4j test application interface. The main content area displays a post titled ". L'identité virtuelle, ma vie au numérique". Below the title, it says: "Afin de prévenir les comportements à risques et de sensibiliser les élèves à l'identité numérique avant leur entrée dans le monde professionnel, cette formation destination des Terminales a été organisée...". A bullet point lists ". NIVEAU ÉDUCATIF DÉTAILLÉ" and "terminale générale et technologique". At the bottom of the post, there is a link "(EDUBASE_391)". On the left side, there is a sidebar with a search bar and a section titled "Base de données Neo4j" which describes the system as being built from metadatabases of archives Edubase collected by the protocol GAI-PMH.



Conclusion

- ❖ Grâce à ce stage, j'ai acquis une compréhension de base et une expérience pratique du concept et du fonctionnement de l'ontologie, ce qui a élargi mes perspectives de carrière. Non seulement j'ai profité de cette occasion pour apprendre à extraire du contenu pertinent via XML et d'autres types de fichiers, mais j'ai aussi appris à exploiter et à créer des bases de données Neo4j et à créer des pages Web via Flask. Cela m'a beaucoup aidé à développer mes compétences professionnelles pour mon avenir.